Spectral Magnitude Modelling for Sinusoidal Coding

Ganan Nagaratnam and David Rowe

Department of Electrical and Computer Engineering
University of Wollongong, Wollongong 2522
AUSTRALIA
email nagar@snrc.uow.edu.au

Abstract

The sinusoidal coder [1] has been shown to be capable of high quality speech at bit rates of 2.4 to 8 kbit/s. This paper addresses the problem of modelling the spectral magnitude parameters for the purposes of low bit rate transmission. Previous work has attempted to fit high order (18-32 pole) LPC models to the spectral magnitudes. This paper explores the use of three hybrid techniques, which use a 12th order all pole model combined with other methods. Two of the schemes presented combine the all pole model with models containing zeros, while the third transmits the modelling error for the perceptually important low order harmonics. Preliminary results indicate that the third method provides near transparent modelling of the spectral magnitudes.

1 Introduction

Sinusoidal coders represent voiced speech as a weighted sum of sinusoids, the model parameters being the sinusoidal frequencies, magnitudes and phases. For low bit rate applications, the frequencies are usually constrained to be harmonics of a fundamental and the phases are usually synthesised at the decoder using perceptually motivated models [1]. Thus for low bit rate sinusoidal coding the parameters are the harmonic magnitudes (spectral magnitudes), the fundamental frequency, and usually some form of voicing estimate. The spectral magnitudes consume the greatest portion of the bit rate, thus compact modelling of the spectral magnitudes is an important research issue in sinusoidal coding.

Direct quantisation of spectral magnitudes is possible, but requires a large number of bits. Also, the number of spectral magnitudes is related to the fundamental frequency and therefore changes on a frame by frame basis. This means direct quantisation would require a time varying bit allocation scheme.

2 LPC Modelling of Spectral Magnitudes

Adjacent spectral magnitudes are highly correlated and tend to describe the vocal tract filtering action. Therefore, the Linear Predictive Coding (LPC) model is a good choice for modelling spectral magnitudes. Techniques for LPC analysis are well known, and LPC parameters are easily transformed to Line Spectral Pair (LSP) frequencies for quantisation and transmission. The spectral magnitude samples may be recovered at the decoder by sampling the LPC model at the harmonic frequencies. Thus the time varying spectral magnitudes may be modelled using a fixed number of LPC coefficients.

One of the problems with all pole models is the inability to adequately model zeros that may be present in the spectrum. During our experiments, we determined that low order all-pole LPC models had difficulty in accurately modelling the low frequency (< 200 Hz) end of the spectrum. One of the reasons for this is that the slope of the all pole model spectrum tends to zero at either end of the spectrum. Since typical male speakers have low fundamental frequencies, the fundamental and lower order harmonics are not accurately modelled. This error is particularly significant when coding band limited (e.g. telephone bandwidth) speech. The perceptual effect is an unpleasant excess of low frequency energy for certain male speakers.

2.1 Pole-Zero Modelling

Pole zero modelling was applied to provide a better fit to the spectral magnitudes than all pole modelling. A non-optimal direct form of modelling (the inverse LP method [2]) was used in the interests of computational simplicity.

2.2 Pole-Selective-Zero

Experimental results suggested that the most significant problem of the all-pole modelling scheme is the inability to model low frequency spectral nulls, such as those

associated with telephone bandwidth band pass filtering. To emphasise the accurate modelling of the lower end of the spectrum, pole zero modelling was confined to a selected low frequency portion of the spectrum. Since zeros are only used in modelling a small portion of the spectrum (typically 0 to 500 Hz), the order of the zeros needed for modelling is reduced. This pole-selective-zero scheme was implemented using frequency domain LPC techniques. To model the formants, the all pole portion of the model was applied to the remaining spectrum.

2.3 All-Pole-Residual

The main difficulty with the all-pole LPC model (in particular for low F0 males) is modelling the first few spectral magnitudes. This approach quantises and transmits the residual error between the all pole model spectrum and the original spectral magnitudes. The calculated residual is the log difference between the spectral magnitude and the sampled all-pole LPC model at the harmonic frequencies. Transmitting the first two to three residuals was sufficient to overcome the low frequency artefacts present with the conventional all-pole model.

3 Results

The three methods were compared to a reference 16th order all pole model in informal subjective listening tests. Objective results obtained from the spectral distortion measure are shown in Table 1. Spectral distortion is averaged across the entire test database. To provide a fair comparison, the pole-zero models were constrained to a maximum of 16th order (eg 10 poles and 6 zeros, or 8 poles and 8 zeros). The test material consisted of two male and two female speakers. Figure 1 shows an example of the 12th order all pole and the selective pole zero (total number of poles 12 and zeros 2) LP models for a single frame of a male speaker.

Model	Zeros	Poles	SD (dB)
All Pole	0	16	3.23
All Pole Residual	0	12	4.07
Pole Selective Zero	low end 2 high end 0	low end 2 high end 10	4.41
Pole Zero	10	6	2.62
	8	8	2.49

Table 1: Spectral Distortion Results

Objective results indicate that the pole-zero model obtained the best results. However informal listening tests indicated that with male speakers the pole-selective-zero scheme was the superior to all-pole and conventional pole-zero techniques for a given model order. This suggests that the pole-selective-zero technique models the perceptually significant harmonic magnitudes more accurately. Overall subjective results indicate that a 12th order LPC model combined with transmitting the first two residuals was sufficient for essentially transparent modelling of the spectral magnitudes. Current research is now directed at efficient quantisation of the model parameters.

4 References

[1] T.G. Champion, R.J. McAulay and T.F. Quatieri, "High-Order Allpole Modelling of the Spectral Envelope," *ICASSP-94*, Vol. 1, pp. 529-531, 1994.

[2] John Makhoul, "Linear Prediction: A Tutorial Review," *Proc. of the IEEE*, Vol. 63, No. 4, April 1975.

[3] H. Morikawa, H. Fujisaki, "Adaptive Analysis of Speech Based on a Pole-Zero Representation,", *Proc. of the IEEE*, Vol. ASSP-30, No. 1, February 1982.

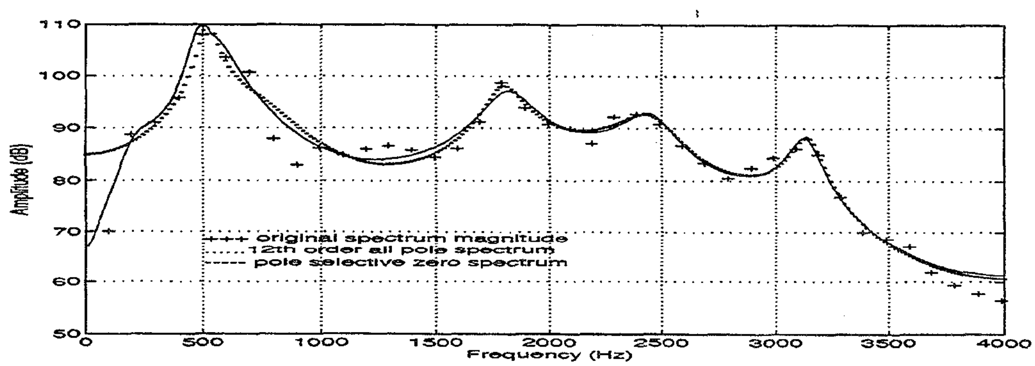

Figure 1. LP Modelling of Spectral Magnitudes